



Resumptive pronouns facilitate processing of long-distance relative clause dependencies in second language English

Fred Zenker & Bonnie D. Schwartz*

Abstract. This study provides evidence that resumptive pronouns (RPs) can facilitate the processing of long-distance subject relative clause (RC) dependencies during second language (L2) sentence comprehension, even where they are disallowed in both the first language (L1) and the target language. A test group of 29 L1-Korean L2 learners (L2ers) of English and a control group of 25 native English speakers completed an online self-paced reading task (SPRT) and an offline acceptability judgment task (AJT) designed to test whether RPs reflect Interlanguage grammar representations and/or a strategy to alleviate processing overload. Analysis of the SPRT data from both response times and comprehension question accuracy indicates that RPs assisted the L2ers, but not the native speakers, with dependency resolution in long-distance RCs. For the AJT data, a proficiency effect was observed whereby some lower-proficiency L2ers, but not the higher-proficiency ones or the native speakers, tended to prefer RPs over gaps in long-distance RCs. The implications of these findings and plans for future research are discussed.

Keywords. English, Korean, relative clauses, resumptive pronouns, second language acquisition, syntax

1. Introduction. This study examines resumptive relative clause (RC) dependencies, a type of phenomenon that is often observed in second language (L2) production even when it is ungrammatical in both the first language (L1) and the target language (TL). In resumptive RCs, a pronoun or full NP occupies the gap position in what would otherwise be a filler-gap dependency. The example in (1), which features a resumptive pronoun (RP) in a subject RC, was produced by an L1-Korean L2 learner (L2er) of English even though subject RPs are generally understood to be ungrammatical in both English and Korean.

- (1) * Sometimes I typed an article, a noun, an adjective, or an adverb that I thought **it** would fit into the context.

These types of Interlanguage (IL) phenomena constitute one of the most interesting areas of L2 research because they cannot be traced to either L1 transfer or TL input, thus offering us a chance to learn more about the basic properties of the human language system.

The current study, a pilot experiment for a larger dissertation research project, investigates whether L2 resumption is part of IL grammar representations and/or stems from a processing strategy that helps L2ers manage cognitive load. A series of online and offline tasks is administered to L1-Korean L2ers of English and to native English controls to probe the processing and acceptability of gaps and RPs in short- and long-distance subject RCs.

* For help of various kinds, we would like to thank Kamil Deen, Haerim Hwang, Li ‘Julie’ Jiang, Hyunwoo Kim, Jihyun Kim, Kitaek Kim, Eunsun Lee, William O’Grady, Amy Schafer, Hong Ri Zenker, Jing ‘Crystal’ Zhong, the Bilinski Dissertation Fellowship and *Language Learning* Dissertation Grant programs, the University of Hawai‘i (UH) Language Analysis and Experimentation Labs, the UH Language Acquisition Reading Group, and all the participants. Authors: Fred Zenker, University of Hawai‘i at Mānoa (fzenker@hawaii.edu), Bonnie D. Schwartz, University of Hawai‘i at Mānoa (bds@hawaii.edu).

2. Background. Filler-gap dependencies such as the RC in (2a) involve a constituent (*the filler*) that is associated with a phonetically empty syntactic position (*the gap*) in a different part of the sentence. Resumptive dependencies like the one in (2b) look very similar to their filler-gap counterparts except that a pronoun or full NP occupies the foot of the dependency, which in true resumptive dependencies is always coreferential with the head NP.

- (2) a. the man_i [that I saw _i]
 b. * the man_i [that I saw him_i]

Some languages have grammatical resumption in a variety of syntactic positions and dependency types (e.g., Asudeh, 2004; McCloskey, 2002). For example, Hebrew can optionally take an RP in direct object RCs, as shown in (3).

- (3) ha-ʔiš_i [še-raʔiti (ʔoto_i)]
 the-man that-saw him
 “the man that I saw”

(Shlonsky, 1992: 444, (1))

However, not all languages are amenable to resumption. Keenan and Comrie (1977) were among the first to collect crosslinguistic data on the distribution of gaps and RPs in RC dependencies. Their findings for English, Korean, and Hebrew are shown in Table 1.

Language	SU	DO	IO	OBL	GEN	OCOMP
English	–	–	–	–	–	–
Korean	–	–	–	–	+	0
Hebrew	–	–/+	+	+	+	+

Table 1. Distribution of gaps (–), RPs (+), and unrelativizable positions (0) in single-clause RCs (adapted from Keenan & Comrie, 1977: 93, Table 2)

Based on RC data from about 50 languages, Keenan and Comrie argued that some syntactic positions are more difficult to relativize from than others. The hierarchy they posited, known as the Noun Phrase Accessibility Hierarchy (NPAH), ranges from subjects (easiest) to objects of comparison (hardest), as shown in (4).

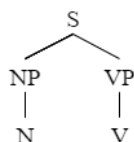
- (4) Subject > Direct Object > Indirect Object > Oblique > Genitive > Object of Comparison

Critical to Keenan and Comrie’s argument is the observation that if a language allows gaps in one position on the hierarchy, it will also allow them in all higher positions. They also noticed that when RPs are permitted at all, they tend to occupy the space between positions that allow gaps and those where relativization is impossible, which they took as evidence that resumptive RCs may be easier in some respects than gap RCs, perhaps because there is an overt category marking the foot of the dependency (p. 92). Depending on the language, RPs may also provide additional information (case, gender, number, etc.) that helps to establish coreference with the head NP.

J. Hawkins (1999) has proposed that the reason some positions are more difficult to relativize from than others has to do with the depth of embedding, which he computed in terms of the size of the filler-gap domain (FGD), i.e., the number of dominating and co-occurring syntactic nodes required by the dependency. As shown in Figure 1 using the accepted syntax of the time, subjects have a minimal FGD of five nodes and direct objects have a minimal FGD of seven nodes. The further we move down the NPAH, the larger the FGD becomes. Other methods of increasing the size of the FGD, such as extracting from embedded complement clauses, can

also increase the amount of processing difficulty. This argument is supported by the existing experimental literature on human sentence processing in native speakers, which has shown that dependencies involving deeper embedding incur higher processing costs, both within clauses (e.g., Just & Carpenter, 1993) and across clause boundaries (e.g., Frazier & Clifton, 1989).

Minimal Domain for Subjects



Minimal Domain for Direct Objects

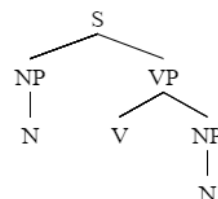


Figure 1. Syntactic trees showing minimal filler-gap domains for subjects and direct objects

While the distribution of gaps and RPs in Korean shown in Table 1 aligns very well with the judgments of our Korean language consultants, it is not universally accepted, and there has been considerable debate among Korean specialists about which positions allow RPs. Kwon (2008) claimed, albeit without experimental evidence, that resumption is possible in both the OBL and GEN positions for single-clause RCs and in all relativizable positions for embedded RCs. By contrast, Song (2003) maintained that Korean allows resumption in only a specific subset of genitive RCs. In addition, Han (2013) provided experimental evidence that Korean speakers tend to reject RPs in subject and direct object RCs formed from simple and (single-level) embedded clauses as well as syntactic islands. These ongoing disagreements underscore the need for further experimental research on RC resumption in Korean.

In fact, the question of whether English might allow RPs in certain types of long-distance dependencies has also been debated. Some researchers, such as Ross (1967), claimed that resumption is “perfectly grammatical” (p. 432) in certain environments that are inaccessible to gaps (i.e., syntactic islands), as shown in (5).

- (5) I just saw [that girl_i who [Long John’s claim [that *(she_i) was a Venusian]] made all the headlines].

(adapted from Ross 1967: 432, (6.154a))

However, recent experimental studies have shown that L1-English speakers consistently assign low ratings to resumptive dependencies both in non-islands and in islands, which casts doubt on claims that English allows RPs in these environments (e.g., Alexopoulou & Keller, 2007; Han et al., 2012; Heestand, Xiang, & Polinsky, 2011; Keffala & Goodall, 2011; McDaniel & Cowart, 1999). Even if resumptive RCs formed from islands are ungrammatical in English, though, it is undeniable that they do occur in the spontaneous speech of native speakers (Asudeh, 2004; Cann, Kaplan, & Kempson, 2005; Prince, 1990). It has been suggested that such utterances are produced not because they are acceptable but because they make the dependency easier to process (e.g., Asudeh, 2004). This kind of resumption is sometimes called *intrusive resumption* (Sells, 1984) to distinguish it from grammatical resumption in languages like Hebrew.

3. Previous research on the processing of resumption in L1 English. There is a growing body of research showing that intrusive resumption, despite being unacceptable, can still facilitate the production and comprehension of difficult-to-process RCs in L1 English (e.g., Beltrama & Xiang, 2016; Ferreira & Swets, 2005; Hammerly, 2020; Hofmeister & Norcliffe, 2013).

One of the earliest of these studies was by Ferreira and Swets (2005) involving a pair of production tasks, one without time pressure and the other with time pressure. Each task consisted of 12 critical items eliciting relativization from positions in *wh*-islands, 12 control items eliciting relativization from non-islands, and 24 fillers. Participants produced resumptive RCs (e.g., “This is a donkey that I don’t know where **it** lives.”) at high rates for critical trials in both experiments: 67% for the untimed experiment and 56% for the timed experiment. The rate at which filler-gap RCs were produced on critical trials was not reported, but it could not have been more than 15% for the untimed experiment and 23% for the timed experiment. The remaining responses consisted of so-called avoidance strategies (e.g., “This is a donkey and I don’t know where it lives.”). In a follow-up offline judgment task, participants tended to reject [island, resumptive] trials and to accept [non-island, gap] trials. These results show that even though RPs are highly unacceptable in English, the resumptive strategy is still preferable to the filler-gap strategy in processing during production when speakers are induced to relativize from islands.

Hofmeister and Norcliffe (2013) used a self-paced reading task to examine how native English speakers process sentences with and without RPs. In the self-paced reading paradigm, slower reading times (RTs) indicate processing difficulty. The study had a 2×2 design crossing *Dependency Length* (short vs. long) and *Resumption* (gap vs. RP), as shown in (6).

- (6) a. Short: The prison officials had acknowledged that there was a prisoner that the guard helped ___/*him to make a daring escape.
 b. Long: Mary confirmed that there was a prisoner who the prison officials had acknowledged that the guard helped ___/*him to make a daring escape.

The results showed that RTs at the region containing the RP were significantly faster in the long environment than in the short environment; RTs at the critical region (the two words immediately following the gap or RP) were also significantly faster in the [long, RP] condition than in any of the other conditions. Taken together, the findings at these two regions suggest that resumption facilitates the comprehension of long-distance RCs, at least in the types of sentences used in the study. However, one important weakness in the Hofmeister and Norcliffe study is that they did not test for correct dependency resolution; we thus cannot be sure that participants associated RPs with the head NP. This problem could be addressed with comprehension questions asking who did what to whom in the sentence (see Morgan, von der Malsburg, Ferreira, & Wittenberg, 2020). Although the task did include comprehension questions, Hofmeister and Norcliffe did not report their content or the by-condition accuracy rates.

The results from these studies point to a processing advantage for RPs over gaps in the production and comprehension of difficult-to-process RCs in English. The current project uses similar methods to test for similar effects in the L2 English of L1-Korean speakers.

4. Previous research on resumption in an L2. Experimental research on L2 resumption has shown that rates of L2 production and acceptance of resumptive RCs tend to be higher in positions thought to be difficult to relativize from (e.g., Algady, 2013; Eckman, Bell, & Nelson, 1988; Gass, 1979; Hyltenstam, 1984; Kim, 2013; Pavesi, 1986). Some studies have also shown that low-proficiency L2ers tend to produce and/or accept RPs at higher rates than high-proficiency ones (e.g., R. Hawkins & Chan, 1997; Kim, 2013; Maghrabi, 1997). Most interesting for the current project, however, are studies showing that L2ers produce resumptive RCs regardless of their grammaticality status in the L1 or the TL (e.g., Gass, 1979; Hyltenstam, 1984; Pavesi, 1986).

Hyltenstam (1984) was the first to demonstrate that L2ers systematically produce resumptive RCs even in environments where they are ungrammatical in both the L1 and the TL. Forty-five L2ers of Swedish from four different L1 backgrounds, only some of which have grammatical resumption in RCs, took part in an oral RC elicitation task with items targeting all six positions on the NPAH. The key finding of the study was that at least some L2ers in each L1 group made regular use of resumption in their production of Swedish RCs, and they were more likely to use RPs in lower positions on the hierarchy than in higher ones. Based on these data, Hyltenstam speculated that L2 resumption arises in the IL from a strategy for reducing cognitive load when producing difficult-to-process RC dependencies.

For Hyltenstam, although the emergence of resumptive RCs in the L2 is driven by processing pressures, this resumption is incorporated into the IL grammar and becomes a licit means of constructing RCs, at least temporarily, at certain levels on the NPAH. While this is indeed a possibility, it is by no means the only one. It is equally possible that resumptive RCs are ungrammatical in the L2ers' IL but they still produce them from time to time as a means of reducing processing load (see also Schulz, 2006, 2011), just as recent experimental studies have suggested is the case in L1 English. If L2ers make more frequent use of the resumptive strategy than their native speaker counterparts, this could be due to the fact that sentence processing in general requires more effort in an L2 than in an L1 (e.g., Kilborn, 1992).

While the idea that L2 resumptive RCs are in some way linked to processing considerations seems plausible enough, it has not yet undergone rigorous experimental testing. Also, there is no study investigating L2 resumption in long-distance RCs, despite research (Alexopoulou & Keller, 2007; J. Hawkins, 1999, 2004; Hofmeister & Norcliffe, 2013) showing that they are more difficult for native speakers to process than the short-distance RCs used in most L2 studies. The current study aims to address these gaps in the literature by administering both online and offline tasks to L1 and L2 speakers of English to investigate the processing and acceptability of gaps vs. RPs in short- and long-distance RC dependencies. This should enable us to disentangle grammatical representations from processing strategies to better understand the status of resumptive RCs in L2 English.

5. Current study. This study uses a series of online and offline tasks to examine the processing and acceptability of gaps vs. RPs in short- and long-distance subject RC dependencies (subject RCs being the environment where RPs are most often observed in English; see Radford, 2019). The test group consists of L1-Korean L2ers of English (Korean is a language that, like English, is widely believed to disallow RPs in simple and embedded subject RCs); native English speakers are also included as controls. The results should help us determine whether L2 resumption reflects IL representations or an ungrammatical processing strategy.

The main tasks include (a) a self-paced reading task (SPRT) designed to test the online processing of gaps vs. RPs during sentence comprehension and (b) an acceptability judgment task (AJT) designed to test the offline acceptability of gaps vs. RPs. Most of the participants also took part in an elicited production task designed to test the processing of gaps vs. RPs during sentence production (the results of which are not reported here). The SPRT was administered before the AJT to minimize the chance of participants consciously evaluating the self-paced reading stimuli for acceptability. The lexical items used in the stimuli for all the tasks were drawn from the official list of 2,315 basic English vocabulary words published by the Korean Ministry of Education, Science, and Technology (MEST; 2008) and from the top 5,000 lemmas from the Corpus of Contemporary American English (Davies, 2008–), which is available at <https://www.wordfrequency.info/freeList.asp>.

If in the SPRT participants have slower RTs at the critical region for gap trials than for RP trials in one or more of the environments tested (and have lower accuracy on the comprehension questions), it could mean that either (a) RPs are easier to process than gaps in the environment(s) or (b) their grammars allow RPs and prohibit gaps there. However, if in the AJT the same participants consistently assign low ratings to RPs in one or more environments, we can take that as evidence that their grammars do not permit resumption there. Triangulating the results from multiple tasks thus allows us to ascertain whether resumptive RCs are a licensed representation for RCs in the IL grammar or the product of an ungrammatical processing strategy.

A summary of conceivable outcomes is shown in Table 2. One possibility is that RPs facilitate processing despite being strongly unacceptable for the L2ers, which would indicate that they represent an ungrammatical processing strategy, just like in L1 English. However, if L2ers systematically accept RPs in one or more conditions on the offline judgment task, it would suggest that resumption is an acceptable option for relativization in their IL grammar.

#	Processing	Representation	Evaluation
1	+	–	intrusive resumption
2	–	+	grammatical resumption
3	+	+	grammatical resumption

Table 2. Possible evaluations of L2-English resumption

Participants also completed a language background questionnaire (at the start of the session) and a 50-item C-test (Zenker, in prep.) to measure English proficiency (at the end of the session), both in-house made. C-tests (Raatz & Klein-Braley, 1981), like cloze tests, are often used by language researchers as a measure of general language ability. In a C-test, the second half of every second word in a text is deleted (excluding repeated words, one-letter words, etc.), and participants are required to fill in the missing letters. There is usually only one correct answer.

The tasks are presented to participants over the internet using Ibex Farm (Drummond, 2007). All data analyses are performed in R (R Core Team, 2019), and all mixed-effects modeling is done using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015).

6. Participants. Twenty-nine L1-Korean L2ers of English and 25 native English speakers participated in this pilot study; their ages, C-test proficiency scores, ages of onset for acquiring English, and years of residence in English-speaking countries are provided in Table 3.

Group	<i>n</i>	Age at Testing	C-test Score	Age of Onset	Years of Residence
L1-Korean	29	26 (19–40)	82 (36–98)	8 (3–19)	2 (0–19)
L1-English	25	21 (18–31)	89 (74–100)	—	—

Table 3. Participant information (values are means and ranges)

The fact that the bulk of the L2ers performed within the native speaker range on the C-test suggests that they were highly proficient learners of English (see Figure 2).

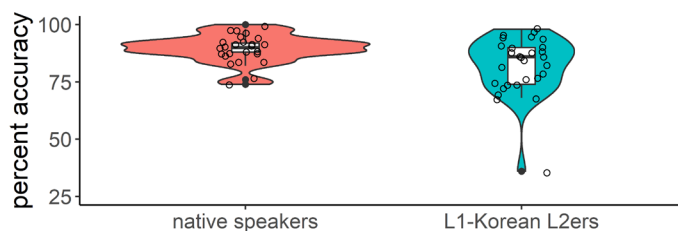


Figure 2. Distribution of proficiency scores from the C-test

7. Self-paced reading task. The SPRT tests whether RPs facilitate RC comprehension. The 24 critical items are distributed across four conditions (see Table 4) in a 2×2 design crossing *Environment* (short vs. long) and *Dependency* (gap vs. RP), the assumption being that long-distance dependencies are harder to process than short-distance ones (e.g., J. Hawkins, 1999). The remaining 36 items are fillers (half grammatical). Each trial is followed by a two-choice comprehension question, with different questions targeting different parts of the sentence; for critical trials, arriving at the correct answer depends on accurate resolution of the RC dependency (e.g., “Who won the race, Mary or the boy?” for the item in Table 4).

Environment	Dependency	Region											
		1	2	3	4	5	6	7	8	9	10	11	12
short	gap	Mary	thinks	that	is	the	boy	that	—	will	win	the	race
short	RP	Mary	thinks	that	is	the	boy	that	he	will	win	the	race
long	gap	That	is	the	boy	that	Mary	thinks	—	will	win	the	race
long	RP	That	is	the	boy	that	Mary	thinks	he	will	win	the	race

Table 4. Conditions for the self-paced reading task (critical region shaded)

Sentences are presented in moving-windows format with word-by-word segmentation, and participants press the spacebar to advance from one word to the next. If resumption facilitates comprehension, RTs should be slower in the [long, gap] condition than in the [long, RP] condition at the critical region, which, following Hofmeister & Norcliffe (2013), consists of the two words following the gap or RP. No differences are expected between the [short, gap] and [short, RP] conditions due to the relative ease of processing short-distance RCs, even for L2ers. Higher comprehension question accuracy for RP trials than for gap trials in (at least) the long environment is also expected if resumption makes RCs easier to process.

Following standard data cleaning procedures for SPRTs (see Jegerski, 2014), all RTs slower than 3,000 ms or faster than 200 ms are excluded from analysis. For each participant group, RTs more than two standard deviations (SDs) above the mean for any region × condition combination are replaced with that group’s cutoff value. A log transformation is also performed on the RTs to make the distribution more normal prior to analysis with linear mixed-effects models.

7.1. RESULTS. All participants had > 80% accuracy on the SPRT comprehension questions (averaged across critical and distractor trials together), indicating that they paid attention while reading the sentences. Visual inspection of the RT data suggested no large differences across conditions for native speakers at the critical region (see Figure 3). By contrast, the L2ers had much slower RTs for the [long, gap] condition than for any other condition at the same region.

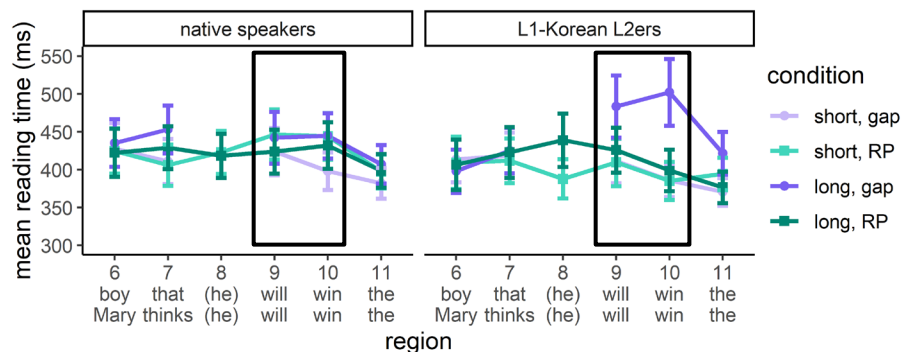


Figure 3. Mean reading times from the self-paced reading task; error bars are 95% confidence intervals; boxes indicate the critical region

The RTs at the critical region were further analyzed for each group using linear mixed-effects regression models entered into R using the following formula: $\text{LogRT} \sim \text{Environment} * \text{Dependency} + (1 + \text{Environment} * \text{Dependency} | \text{Participant}) + (1 + \text{Environment} * \text{Dependency} | \text{Item})$. For the native speakers, there were no significant effects for *Environment* or *Dependency*, but the interaction term was significant ($\beta = 0.185, p = .008$). Planned pairwise comparisons showed that RP trials were read significantly slower than gap trials in the short environment ($\beta = 0.133, p = .014$) but not in the long environment; this indicates, if anything, that RPs made processing slightly harder for the native speakers in the short environment, perhaps due to surprisal. For the L2ers, the full-model results revealed that there was a significant effect for *Environment* ($\beta = 0.112, p = .002$), a marginally significant effect for *Dependency* ($\beta = 0.071, p = .061$), and a significant interaction between the two ($\beta = 0.137, p = .041$). Planned pairwise comparisons showed gap trials were read significantly slower than RP trials in the long environment ($\beta = 0.137, p = .002$) but not in the short environment, thus indicating that RPs were easier for the L2ers to process than gaps in only the long environment.

Figure 4 shows that for both groups, comprehension question accuracy rates were numerically higher for RP trials than for gap trials in (at least) the long environment.

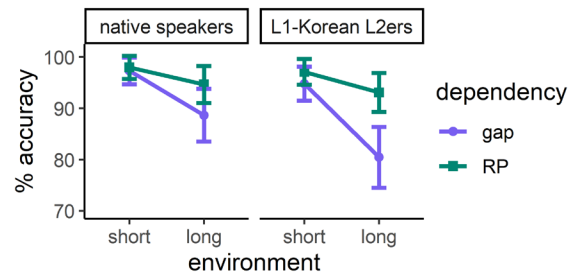


Figure 4. Mean comprehension question accuracy rates from the self-paced reading task; error bars are 95% confidence intervals

The accuracy data for each group were examined further by means of logistic mixed-effects regression models using the following formula: $\text{Accuracy} \sim \text{Environment} * \text{Dependency} + (1 + \text{Environment} * \text{Dependency} | \text{Participant}) + (1 + \text{Environment} * \text{Dependency} | \text{Item})$. No significant effects were found for the variables or their interactions for either group, a result which is attributable to the fact that only a subset of the participants had difficulty understanding sentences in the [long, gap] condition. Subsequent inspection of the individual participant response data revealed that the pattern of results shown in Figure 4—where accuracy rates are higher for RP trials than for gap trials in (at least) the long environment—was driven by about a third of the participants in each group (36% for native speakers and 34% for L2ers). The results from this descriptive statistics analysis show that at least some of the participants in both groups found it easier in the long environment to understand sentences with RPs than those with gaps, although the magnitude of this difference was larger for L2ers than for native speakers.

7.2. DISCUSSION. Taken together, the results for the RT data and the comprehension question accuracy data from the L2ers both point to a processing effect for RPs in long-distance RC dependencies. This effect indicates (a) that these L2 participants experienced at least some difficulty parsing sentences in the long environment and (b) that the presence of an RP assisted with dependency resolution. For the native speakers, there was no evidence of processing facilitation from RPs in the RT data, and the magnitude of the effect observed in their comprehension question accuracy data was smaller than that of the L2ers. This suggests that the

sentences used in this study were not complex enough (cf. sentences used in the long condition of Hofmeister and Norcliffe, 2013, exemplified in (6) above) to cause serious processing difficulty for the native speaker participants (as we might expect, seeing as these sentence types are relatively short, simple, and frequent).

8. Acceptability judgment task. The AJT tests the acceptability of the sentence types in the SPRT; the two tasks have the same design, but different lexicalizations are used for the stimuli. Participants rate sentences on a 6-point Likert scale with an additional I-don't-know option. Following standard data analysis practices for AJTs, responses are converted to z-scores to minimize scale bias (e.g., Sprouse, Wagers, & Phillips, 2012).

8.1. RESULTS. Visual inspection of the mean z-score ratings shows that the participants in both groups tended to rate gap trials higher than RP trials across the two environments (see Figure 5).

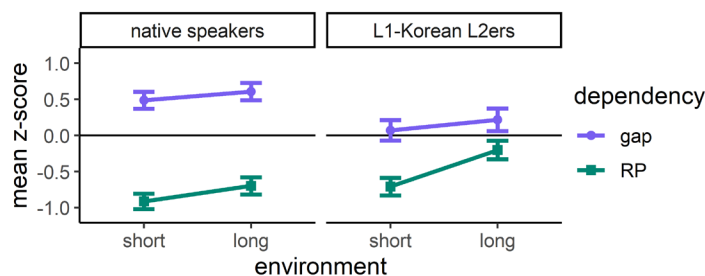


Figure 5. Mean z-score ratings from the acceptability judgment task; error bars are 95% confidence intervals

The z-score ratings for each group were further examined with linear mixed-effects regression models using the following formula: $\text{Rating} \sim \text{Environment} * \text{Dependency} + (1 + \text{Environment} * \text{Dependency} | \text{Participant}) + (1 + \text{Environment} * \text{Dependency} | \text{Item})$. For the native speakers, there was a marginally significant effect for *Environment* with the long trials rated higher than short trials ($\beta = 0.168, p = .057$) and a significant effect for *Dependency* with gaps rated higher than RPs ($\beta = 1.360, p < .001$), but the interaction term was not significant. The L2ers performed very similarly to the native speakers when analyzed as a group; there were significant effects for *Environment* ($\beta = 0.327, p = .037$) and *Dependency* ($\beta = 0.598, p = .002$), but not for their interaction. These results confirm our earlier observation that both groups tended to accept gaps and reject RPs across the two environments.

To investigate whether individual differences in proficiency influenced how the L2ers judged gaps and RPs in the long environment, a simple linear regression analysis was performed to compare their difference scores (i.e., the mean z-score for the [long, gap] condition minus the mean z-score for the [long, RP] condition) to their C-test proficiency scores (see Figure 6).

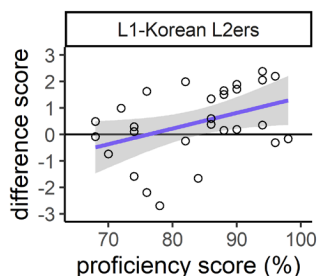


Figure 6. Difference scores plotted against proficiency scores for the L1-Korean L2ers on the acceptability judgment task

One extreme outlier whose proficiency score was more than three SDs below the mean was removed from this analysis. The remaining results showed that there was a significant relationship between the two sets of values ($F = 4.98$, $R^2 = .16$, $p = .034$) stemming from the fact that lower-proficiency L2ers tended to have negative difference scores (indicating that they preferred RPs to gaps in the long environment) and higher-proficiency L2ers tended to have positive difference scores (indicating that they preferred gaps to RPs in the same environment).

8.2. DISCUSSION. The native speakers performed as expected on the AJT, assigning high ratings to sentences with gaps and low ratings to those with RPs regardless of dependency length, indicating that their grammars do not permit resumption in either environment. The L2ers had the same general pattern of results as a group, but proficiency effects were also detected which suggest that some low-proficiency L2ers accept RPs in long-distance RCs.

One unexpected finding was that both the L1 and L2 groups gave numerically lower ratings for gap trials in the short environment than in the long environment; we would expect the opposite if long-distance RCs are more difficult to process. A first stab at understanding this was that it might have resulted from a garden-path effect, i.e., the *that* in *Mary thinks that...* initially parsed as a complementizer (see Table 4). However, data from an additional eight native English speakers with *that* changed to *this* across all trials to eliminate potential garden-pathing simply replicated the same general rating pattern, as shown in Figure 7. It thus remains unclear exactly why ratings for gaps were slightly lower in the short environment than in the long environment.

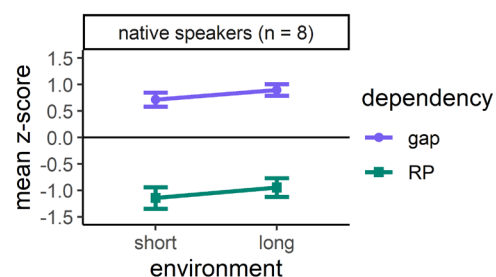


Figure 7. Mean z-scores on an updated version of the acceptability judgment task with *that* changed to *this* across all stimuli. Error bars are 95% confidence intervals

9. General discussion. This study tested whether L2 English resumption is a licit option for subject relativization in IL grammars and/or facilitates the processing of them. For the native speaker controls, the SPRT data provided little or no evidence that RPs ease comprehension in either short- or long-distance subject RCs. This result was expected because the sentences were relatively short and simple, even in the long-distance environment. If we made the dependencies harder to resolve, either by adding additional layers of embedding or by placing the foot of the dependency inside a syntactic island, then we might very well uncover processing effects even for the native speakers (see Hammerly, 2020; Hofmeister & Norcliffe, 2013).

On the AJT, the same native speaker participants gave consistently low ratings for RPs in the two RP conditions. The fact that RP trials were consistently rejected in both the short and long environments suggests that resumption is not a grammatical means of forming RCs for these participants, at least not for the sentence types used in the stimuli. This finding is very much in line with previous experimental research with native English speakers indicating that RPs are ungrammatical in short-distance, long-distance, and island RCs (e.g., Han et al., 2012; Heestand et al., 2011; Keffala & Goodall, 2011; McDaniel & Cowart, 1999).

The L2 data patterned differently from the L1 data in several important ways. For one thing, in the SPRT, both the RT data and the comprehension question accuracy data indicated that at least a non-trivial portion of the L2ers had trouble processing RCs with a gap in an embedded clause and that the insertion of an RP made it easier for them to resolve this type of dependency. These results are consistent with Hylenstam's (1984) hypothesis that RPs make RC dependencies in an L2 easier to process. The finding that the L2ers had more trouble than the native speakers did with processing the long-distance gap trials also adds to the body of evidence that sentence processing is slower and more effortful in an L2 than in an L1 (e.g., Kilborn, 1992).

As for the AJT data, the L2ers as a group performed similarly to the native speakers, giving fairly low ratings for RPs across conditions. However, a proficiency effect was also observed whereby some of the lower-proficiency L2ers tended to prefer RPs to gaps in long-distance RCs. More data collection is needed to determine whether RPs are part of the IL grammar for low-proficiency L2ers, but the results thus far at least suggest that this is a possibility.

9.1. DISAGREEMENTS SURROUNDING PROCESSING EFFECTS FOR RPs. Although researchers interested in L1 English resumption in RCs tend to agree that RPs can facilitate processing during sentence production (e.g., Ferreira & Swets, 2005; Heestand et al., 2011; Polinsky, Clemens, Morgan, Xiang, & Heestand, 2013), it should be acknowledged that not all previous studies have found that RPs are helpful for comprehension, even in difficult-to-process RC dependencies such as those involving extraction from islands. In fact, Morgan et al.'s (2020) SPRT results showed that RPs can actually hinder comprehension under certain conditions.

What, then, is the critical difference between the studies by Hofmeister and Norcliffe (2013) and Hammerly (2020) on the one hand, which found that RPs ease comprehension, and the one by Morgan et al. (2020) on the other hand, which found that RPs hamper comprehension? The answer could lie in the amount of useful information that the RP provides in the given context. The studies by Hofmeister and Norcliffe and by Hammerly as well as the current study all used stimuli where the RP carries gender information that matches only the head NP in the sentence, thereby providing information that can assist with dependency resolution. In Morgan et al.'s study, by contrast, the person, number, gender, and case marking on the RP matched more than one of the NPs in the sentence—as in the example *It is Mr. Dino that Mr. Rabbit wondered whether Miss Piggy tickled him with a feather* (p. 4, Figure 1)—and thus there was no overt marking on the RP that uniquely identified the head NP.

In light of these observations, perhaps we should not be asking whether resumption always necessarily facilitates comprehension in difficult-to-process RC dependencies but instead whether it is capable of facilitating comprehension under favorable conditions. The current study's findings add to a growing body of research showing that RPs in RCs can facilitate comprehension when (a) the sentences types are complex enough to cause processing difficulty and (b) the RP is marked for phi-features that assist with (unique) dependency resolution.

9.2. LIMITATIONS AND FUTURE DIRECTIONS. This study suffered from several limitations that will need to be addressed in future iterations of the experiment. One of these limitations was low statistical power. Analysis of the current dataset with the *simr* package in R (Green & McLeod, 2016) showed that at least 60 participants per group would be needed to achieve sufficient power ($\geq 80\%$) for detecting an interaction between the *Environment* and *Dependency* factors on the SPRT (the most power-hungry of the tasks), if indeed one exists. This problem can be addressed by increasing the sample sizes for upcoming rounds of data collection.

A second limitation of this study is that there was an insufficient range of proficiency levels among the L2ers. The present dataset therefore cannot give us a very good sense of how proficiency modulates performance on the tasks. For future data collection, it will be important to find more L2ers with lower English proficiency through targeted participant recruitment.

Another similar problem is that there was too much variability in age of onset for the L2ers, with one participant even starting to learn English at the young age of three (however, this participant also reported that she had not spent any time living in English-speaking countries during childhood). As a result, not all the L2ers in this study can be categorized as adult learners. Unless one plans to make comparisons between early and late acquirers, it would be preferable to limit participation to those who started learning English at no younger than, say, age 10.

There are a number of other ways in which the current study was rather limited in scope. For example, there were only L2ers from a single L1 background. Upcoming rounds of data collection will include a group of L1-Mandarin L2ers of English (Mandarin being a language that is generally taken to have grammatical RC resumption in a range of syntactic positions), which will allow for the exploration of L1 effects, at the levels of both grammar and processing. Future data collection will also include experiments probing direct object RCs and a *wh*-island condition, which will allow for more direct comparisons with existing literature on the processing and acceptability of RPs in L1 English.

Finally, the L2ers were tested only in English, which means that we cannot know for certain how they would have performed in their L1. This is an important limitation because the facts regarding the acceptability of resumption in Korean (and Mandarin) are still not well understood (see Keenan & Comrie, 1977; Kwon, 2008; Pan, 2016; Song, 2003), and also because it is possible that there is variation between speakers with the same L1. This problem will be addressed moving forward by testing the L2ers in their L1 as well as the TL, at least for the AJT. This is an important step to include in L2 research of this type, and one that has only very seldom been used in previous L2 studies of any kind (but see Zenker & Schwartz, 2017).

10. Conclusion. This study takes methods from the recent processing literature on resumptive RCs in L1 English and adapts them for investigating the same phenomenon in L2 English. It is the first study that systematically tests Hyldenstam's (1984) hypothesis that L2 resumption in RCs is linked to processing considerations, focusing specifically on processing during sentence comprehension. Both online and offline tasks were administered to L1-Korean L2ers of English—Korean being a language that is generally thought to disallow RPs in both short- and long-distance subject RCs—to explore whether resumption is a licit option for subject relativization in the IL grammar and/or is the result of a strategy for reducing cognitive load during sentence processing. The data indicate that RPs can facilitate L2 sentence processing, at least for difficult-to-process RCs. The data also suggest that the acceptability of RPs is modulated by L2 proficiency. Future iterations of this study will use revised materials (for subject RCs as well as include direct object RCs and an additional *wh*-island condition) and increased sample sizes and will test L2 groups with contrastive L1s in hopes of achieving a deeper understanding of the processing and acceptability of RPs in L2 English RCs.

References

- Alexopoulou, Theodora & Frank Keller. 2007. Locality, cyclicity, and resumption: At the interface between the grammar and the human sentence processor. *Language* 83. 110–160. <https://doi.org/10.1353/lan.2007.0001>.
- Algady, Dola. 2013. *The acquisition of relative clauses: How do second language learners of Arabic do it?* Milwaukee, WI: University of Wisconsin-Milwaukee dissertation.
- Asudeh, Ash. 2004. *Resumption as resource management*. Stanford, CA: Stanford University dissertation.
- Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67. 1–48.
- Beltrama, Andrea & Ming Xiang. 2016. Unacceptable but comprehensible: The facilitation effect of resumptive pronouns. *Glossa* 1. 29.1–24. <https://doi.org/10.5334/gjgl.24>.
- Cann, Ronnie, Tami Kaplan & Ruth Kempson. 2005. Data at the grammar-pragmatics interface: The case of resumptive pronouns in English. *Lingua* 115. 1551–1577. <https://doi.org/10.1016/j.lingua.2004.07.004>.
- Davies, Mark. 2008–. Corpus of Contemporary American English. <https://www.english-corpora.org/coca/>.
- Drummond, Alex. 2007. *Ibex farm*. <http://spellout.net/ibexfarm>.
- Eckman, Fred R., Lawrence Bell & Diane Nelson. 1988. On the generalization of relative clause instruction in the acquisition of English as a second language. *Applied Linguistics* 9. 1–20. <https://doi.org/10.1093/applin/9.1.1>.
- Ferreira, Fernanda & Benjamin Swets. 2005. The production and comprehension of resumptive pronouns in relative clause “island” contexts. In Anne Cutler (ed.), *Twenty-first century psycholinguistics: Four cornerstones*. 263–278. Mahwah, NJ: Lawrence Erlbaum.
- Frazier, Lyn & Charles Clifton. 1989. Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes* 4. 93–126. <https://doi.org/10.1080/01690968908406359>.
- Gass, Susan M. 1979. Language transfer and universal grammatical relations. *Language Learning* 29. 327–344. <https://doi.org/10.1111/j.1467-1770.1979.tb01073.x>.
- Green, Peter & Catriona J. MacLeod. 2016. simr: An R package for power analysis of generalized linear mixed models by simulation, *Methods in Ecology and Evolution* 7. 493–498. <https://doi.org/10.1111/2041-210x.12504>.
- Hammerly, Christopher. 2020. Intrusive resumption can ameliorate island violations in real-time comprehension. Working paper, University of Massachusetts, Amherst. <https://ling.auf.net/lingbuzz/004846>.
- Han, Chung-hye. 2013. On the syntax of relative clauses in Korean. *Canadian Journal of Linguistics* 58. 319–347.
- Han, Chung-hye, Noureddine Elouazizi, Christina Galeano, Emrah Görgülü, Nancy Hedberg, Jennifer Hinnell, Meghan Jeffrey, Kyeong-min Kim & Susannah Kirby. 2012. Processing strategies and resumptive pronouns in English. In Nathan Arnett & Ryan Bennett (eds.), *Proceedings of the 30th West Coast Conference on Formal Linguistics*, 153–161. Somerville, MA: Cascadilla Press. <http://www.lingref.com/cpp/wccfl/30/abstract2813.html>.
- Hawkins, John A. 1999. Processing complexity and filler-gap dependencies across grammars. *Language* 75. 244–285. <https://doi.org/10.2307/417261>.
- Hawkins, John A. 2004. *Efficiency and complexity in grammars*. New York: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199252695.001.0001>.

- Hawkins, Roger & Cecilia Yuet-hung Chan. 1997. The partial availability of Universal Grammar in second language acquisition: The 'failed functional features hypothesis.' *Second Language Research* 13. 187–226. <https://doi.org/10.1191/026765897671476153>.
- Heestand, Dustin, Ming Xiang & Maria Polinsky. 2011. Resumption still does not rescue islands. *Linguistic Inquiry* 42. 138–152. https://doi.org/10.1162/ling_a_00032.
- Hofmeister, Philip & Elisabeth Norcliffe. 2013. Does resumption facilitate sentence comprehension? In Philip Hofmeister & Elisabeth Norcliffe (eds.), *The core and the periphery: Data-driven perspectives on syntax inspired by Ivan A. Sag*. 225–246. Stanford, CA: CSLLI Publications.
- Hyltenstam, Kenneth. 1984. The use of typological markedness conditions as predictors in second language acquisition: The case of pronominal copies in relative clauses. In Roger W. Andersen (ed.), *Second languages: A cross-linguistic perspective*. 39–58. Rowley, MA: Newbury House.
- Jegerski, Jill. 2014. Self-paced reading. In Jill Jegerski & Bill VanPatten, *Research methods in second language psycholinguistics*. 20–49. New York: Routledge. <https://doi.org/10.4324/9780203123430>.
- Just, Marcel A. & Patricia A. Carpenter. 1993. The intensity dimension of thought: Pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology* 47. 310–339. <https://doi.org/10.1037/h0078820>.
- Keenan, Edward L. & Bernard Comrie. 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8, 63–99.
- Keffala, Bethany & Grant Goodall. 2011. Do resumptive pronouns ever rescue illicit gaps in English? Poster presented at the 2011 CUNY Conference on Human Sentence Processing, Stanford University, Stanford, CA.
- Kilborn, Kerry. 1992. On-line integration of grammatical information in a second language. In Richard J. Harris (ed.), *Cognitive processing in bilinguals*. 337–350. New York: North-Holland. [https://doi.org/10.1016/s0166-4115\(08\)61504-6](https://doi.org/10.1016/s0166-4115(08)61504-6).
- Kim, Chae-Eun. 2013. *Asymmetries in the production of relative clauses: First and second language acquisition*. Honolulu, HI: University of Hawai'i dissertation.
- Kwon, Nayoung. 2008. *Processing of syntactic and anaphoric gap-filler dependencies in Korean: Evidence from self-paced reading time, ERP and eye-tracking experiments*. La Jolla, CA: University of California, San Diego dissertation.
- Maghrabi, Aimen A. 1997. *The roles of psycholinguistic constraints and typological influence in the acquisition of pronominal copies in relativization by Arabic and English learners*. Washington, DC: Georgetown University dissertation.
- McCloskey, James. 2002. Resumption, successive cyclicity, and the locality of operations. In Samuel D. Epstein & T. Daniel Seely (eds.), *Derivation and explanation in the minimalist program*. 184–226. Malden, MA: Blackwell. <https://doi.org/10.1002/9780470755662.ch9>.
- McDaniel, Diana & Wayne Cowart. 1999. Experimental evidence for a minimalist account of English resumptive pronouns. *Cognition* 70. B15–B24. [https://doi.org/10.1016/s0010-0277\(99\)00006-2](https://doi.org/10.1016/s0010-0277(99)00006-2).
- Ministry of Education, Science, and Technology. 2008. The school curriculum of the Republic of Korea: Proclamation of the Ministry of Education, Science, and Technology #2008-160.
- Morgan, Adam M., Titus von der Malsburg, Victor S. Ferreira & Eva Wittenberg. 2020. Shared syntax between comprehension and production: Multi-paradigm evidence that resumptive

- pronouns hinder comprehension. *Cognition* 205. Article 104417.
<https://doi.org/10.1016/j.cognition.2020.104417>.
- Pan, Victor J. 2016. *Resumptivity in Mandarin Chinese: A minimalist account*. Boston: De Gruyter.
- Pavesi, Maria. 1986. Markedness, discursial modes, and relative clause formation in a formal and an informal context. *Studies in Second Language Acquisition* 8. 38–55.
<https://doi.org/10.1017/s0272263100005829>.
- Polinsky, Maria, Lauren E. Clemens, Adam M. Morgan, Ming Xiang & Dustin Heestand. 2013. Resumption in English. In Jon Sprouse & Norbert Hornstein (eds.), *Experimental syntax and island effects*. 341–359. New York: Cambridge University Press.
<https://dash.harvard.edu/handle/1/33954794>.
- Prince, Ellen. F. 1990. Syntax and discourse: A look at resumptive pronouns. In Kira Hall, Jean-Pierre Koenig, Michael Meacham, Sondra Reinman, & Laurel A. Sutton (eds.), *Proceedings of the 16th annual Meeting of the Berkeley Linguistics Society*. 482–497. Berkeley, CA: Berkeley Linguistics Society. <https://doi.org/10.3765/bls.v16i0.1719>.
- R Core Team. 2020. *R: A language and environment for statistical computing*. <https://www.R-project.org>.
- Raatz, Ulrich & Christine Klein-Braley. 1981. The C-test: A modification on the cloze procedure. In Terry Culhane, Christine Klein-Braley & Douglas K. Stevenson (eds.), *Practice and problems in language testing, University of Essex Occasional Papers* 26. 113–138. Colchester, UK: University of Essex.
- Radford, Andrew. 2019. *Relative clauses: Structure and variation in everyday English*. New York: Cambridge University Press. <https://doi.org/10.1017/9781108687744>.
- Ross, John R. 1967. *Constraints on variables in syntax*. Cambridge, MA: Massachusetts Institute of Technology dissertation.
- Schulz, Barbara. 2006. *Wh-scope marking in English Interlanguage grammars: Transfer and processing effects on the second language*. Honolulu, HI: University of Hawai‘i dissertation.
- Schulz, Barbara. 2011. Syntactic creativity in second language English: *Wh-scope marking in Japanese-English interlanguage*. *Second Language Research* 27. 313–341.
<https://doi.org/10.1177/0267658310390503>.
- Sells, Peter. 1984. *Syntax and semantics of resumptive pronouns*. Amherst, MA: University of Massachusetts, Amherst dissertation.
- Shlonsky, Uri. 1992. Resumptive pronouns as a last resort. *Linguistic Inquiry* 23. 443–468.
- Song, Jae Jung. 2003. Resumptive genitive pronouns in Korean relative clauses: Distribution and explanation. *SKY Journal of Linguistics* 16. 139–160.
- Sprouse, Jon, Matt Wagers & Colin Phillips. 2012. A test of the relation between working-memory capacity and syntactic island effects. *Language* 88. 82–123.
<https://doi.org/10.1353/lan.2012.0004>.
- Zenker, Fred. in preparation. *Development and validation of a C-test for the assessment of general English ability*. Manuscript, University of Hawai‘i.
- Zenker, Fred & Bonnie D. Schwartz. 2017. Topicalization from adjuncts in English vs. Chinese vs. Chinese-English Interlanguage. In Maria LaMendola & Jennifer Scott (eds.), *Proceedings of the 41st annual Boston University Conference on Language Development, Vol. 2*. 806–819. Somerville, MA: Cascadia Press.
<http://www.lingref.com/buclid/41/BUCLD41-65.pdf>.